第40卷 第9期
2023年9月

干 旱 区 研 究
ARID ZONE RESEARCH

Vol. 40 No. 9
Sep. 2023

# 深度学习方法下GEDI数据的天然云杉林地上生物量反演

孙丹阳[1,2]， 魏建新[1,2,3]， 杨 辽[4]， 王 杰[5]， 唐宇琪[4]， 巴比尔江·迪力夏提[2,3]

（1. 新疆大学地理与遥感科学学院，新疆 乌鲁木齐 830017；2. 新疆激光雷达应用工程技术研究中心，新疆 乌鲁木齐 830002；3. 新疆维吾尔自治区自然资源信息中心，新疆 乌鲁木齐 830002；4. 中国科学院新疆生态与地理研究所，荒漠与绿洲生态国家重点实验室，新疆 乌鲁木齐 830011；5. 西华师范大学地理科学学院，四川 南充 637002）

**摘 要**：森林作为陆地最大碳库，对人类的生活与发展至关重要，精准掌握森林资源动态变化并对其进行现代化可持续发展已成为当下研究热点。本文以天山山脉的天然云杉林为研究对象，利用地面实测数据、直升机机载激光雷达点云数据以及全球生态系统动态调查激光雷达（Global Ecosystem Dynamics Investigation，GEDI）数据，构建多源融合数据框架，通过使用AutoKeras框架下的深度学习算法，实现GEDI数据的多个相对高度百分位数（Relative Height Percentile，RH）与其光斑内地上生物量的回归模型预测，验证GEDI数据在较大范围的地上生物量反演方面的可行性，主要结论如下：（1）GEDI数据用于森林地上生物量估测研究具有较高可行性，通过自动化深度学习算法，训练集、验证集、整体数据的决定系数（Coefficient of Determination，$R^2$）分别为0.69、0.63和0.67，平均绝对误差（Mean Absolute Error，MAE）分为3.73 mg·hm$^{-2}$、4.22 mg·hm$^{-2}$和3.89 mg·hm$^{-2}$，具有较高的预测精度。（2）直升机激光雷达作为GEDI数据估算地上生物量的中间技术，整个研究区内的单木识别准确率高于0.75。最终本次研究通过多模态数据融合，定量化描述研究区单木基础结构参数的同时，验证GEDI数据在获取森林地上生物量方面的潜力，也为相近区域大面积的森林碳源汇、生物量、蓄积量估算、森林管理与经营、生物多样性保护等多个项目研究提供理论基础，具有一定的指导意义和基础数据支撑作用。

**关键词**：天然云杉林；GEDI；LiDAR；地上生物量；深度学习

森林生态系统是全球生态环境的重要组成部分，是释放和吸收二氧化碳的主要碳库，研究表明，森林是陆地生态系统中最大的有机碳储库，占全球总植被碳储量的86%[1]。因此，准确、高效的获得森林单木胸径、树高、冠幅等结构因子至关重要，可为生物量、碳储量求解提供数据支撑，为可持续森林管理提供科学依据。传统的人工野外森林资源普查，受环境限制较大，操作复杂、数据精度容易受人为主观影响无法满足现代林业资源调查、检测等活动。

激光雷达（Light Detection And Ranging，LiDAR）作为一项主动遥感技术，具有数据密度大、分辨率高、穿透性强、数据精度高、受天气影响小等技术优势[2-3]，可以无接触、低成本、无破坏的获得森林结构参数和林下地形信息。各类型激光雷达技术相互补充，构建多源数据融合框架，以便精确获得树冠内部和林下地形的空间特征信息[4]，为估算大范围的地上生物量（Above-Ground Biomass，AGB）、蓄积量等提供数据支撑。其中星载激光雷达主要以卫星为搭载平台，可以用于全球尺度的地表特征和变化检测，主要应用于全球测绘、地球科学、大气探测等活动[5]。2014年马里兰大学和美国航天航空局（National Aeronautics and Space Administration，NASA）戈达德太空飞行中心合作，建造并测试了全球生态系统动态调查（Global Ecosystem Dynamics Investigation，GEDI）高分辨率激光测距系统，并于

**收稿日期**：2023-04-01； **修订日期**：2023-05-16
**基金项目**：新疆维吾尔自治区财政拨款项目（65000022P00790610001Q）
**作者简介**：孙丹阳（1998-），女，硕士研究生，研究方向为地信工程. E-mail: 1063209224@qq.com
**通讯作者**：魏建新. E-mail: wjxlr@126.com

2018年在国际空间站部署[6],是第一个专门设计用于获取植被结构的星载激光雷达系统[7],主要应用于南北纬51.6°之间热带和温带地区的森林冠层高度、林下地形和地上生物量等精准测量[8],致力于提高气候和土地利用变化对生态系统结构和动态影响的表征能力。

从研究方法、技术手段方面可以发现,人工智能和多来源、多时间、多数量的林业数据使用频率日益增加,传统的森林资源线性回归需要满足一定的条件假设,深度学习已成为机器学习和人工智能的一种突出方法[9],国内外学者在深度学习、定量遥感获取森林参数方面进行广泛研究。Cao等[10]总结生物量提取方法时,通过多元线性回归、神经网络等方法构建多源遥感参数与生物量的关系;Torre-Tojal等[11]使用随机森林模型估计森林生物量,采用2折交叉验证获得模型拟合精度$R^2$高于0.7;Quirós等[12]选取GEDI的RH100作为林分垂直结构,与LiDAR比较,地面高程差的RMSE为6.13 m;Adam等[7]发现GEDI与机载LiDAR导出的数字地形模型具有高相关性($R^2$=0.99),而森林冠层模型之间则显示相当弱的线性关系($R^2$=0.27);Fayad等[13]则对林分优势高度采用随机森林回归建模方法对其中重要性变量进行排序,GEDI的RH70、RH80、RH90、RH100占前四,可以发现GEDI数据对生态系统动态监测等活动具有重要指导意义[7,14-15],同时国内外学者对于GEDI数据基础参数的属性设置方面也进行了大量研究与探讨。

Fayad等[16]利用GEDI数据估算瑞士8个湖泊的水位精度,虽然与现场水位吻合效果较好,但是处于早晨或是夜晚采集的GEDI数据要比白天或是周末的数据具有更低的偏离。Potapov等[17]也选择了在夜间以功率激光器运行收集,且光束灵敏度为90%的GEDI数据,同时对地面高程的范围也进行设置,在该研究中,除去过高(>30 m)和过矮(<7 m)的冠层高度值,GDEI的RH95与基于Landsat的森林高度图有良好的对照关系。林晓娟[18]则直接使用GEDI的光斑高程,选择RH50、RH75、RH90、RH95、RH100与机载LiDAR 30 m空间分辨率的森林冠层高度栅格产品进行比对,选择最能代表森林冠层高度的属性字段,在增加GEDI的7个筛选条件之后,两者之间的$R^2$明显增大,各类误差逐渐减小,最终选择RH95(平均误差为1.43 m)代表GEDI光斑的冠层高度。朱笑笑[19]发现GEDI数据提取的RH95森林高度与机载LiDAR数据提取的森林高度相关性最高($R^2$=0.95,RMSE=2.22),相比于ICESat-2/ATLAS,GEDI提取的森林高度精度更高。Ngo等[20]则检验了RH指标从90%~100%的性能,验证GEDI数据的RH98是研究CHM的一个很好的性能指标,事实上,RH98也是一个较为常见的参数选择。

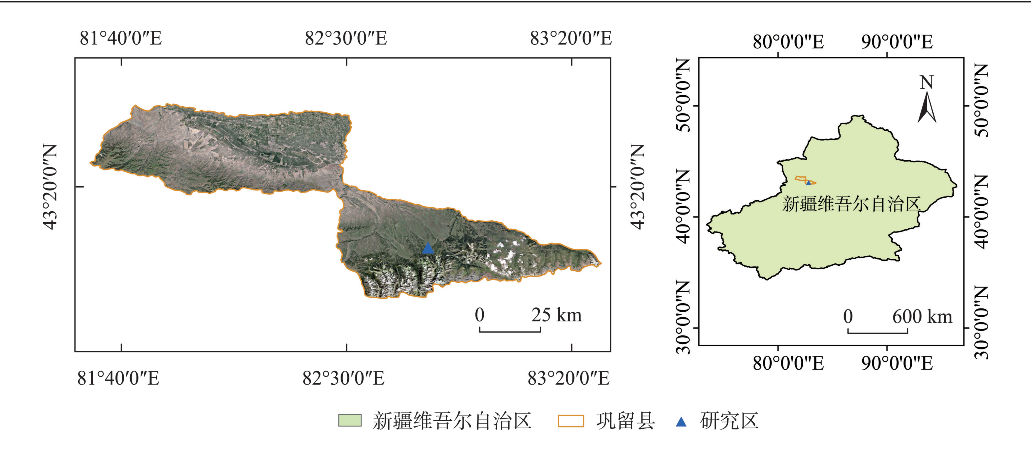
根据以上分析可以发现GEDI中RH值虽然作为森林高度代表,但不是一个确定的测量方法,因此对于GEDI数据反演各类森林参数时通常选择RH95或是单独的几个具有代表性的相对高度百分位数作为反演GEDI光斑内的森林冠层高度的重要参数[17,21-23],同时可以发现构建多源数据融合框架,快速高效估算生物量等其他森林功能参数成为必然趋势,利用其他影像数据提取地形、坡度、归一化植被指数等因子可以有效提高模型预测精度。对于GEDI数据的参数使用方面,通常使用有限个RH值作为GEDI数据的光斑尺度森林高度值,例如,使用的RH100代表光斑内的最高树冠顶层,这种方法虽然便捷高效,但忽略了光斑内的植被分布特征,未考虑所有RH值对于最终计算的区域性森林功能参数的影响因素,存在一定的低估现象。

因此,本文结合样方实测单木数据、机载LiDAR点云数据、GEDI数据,构建多模态数据融合,以GEDI点内81个非负RH值($RH_{20}$~$RH_{100}$,1%间隔)为输入变量,AGB为输出变量,尽可能多的保留GEDI光斑内的植被分布特征,使用基于AutoKeras框架的深度学习方法构建AGB回归模型,减少调参与模型构建压力的同时,遍历更多模型,提高模型预测精度,高精度的实现大范围的地上生物量估算,验证GEDI在实现AGB估算方面的潜力,更为区域性AGB估算提供方法依据和技术借鉴。

## 1 研究区概况与数据处理

### 1.1 研究区概况

研究区位于新疆巩留县,地理位置位于81°34′~83°35′E,42°54′~43°58′N(图1),隶属于伊犁哈萨克自治州,东南与和静县相邻,主要由山地、丘陵、平原三大地貌单元组成,海拔在630~4230 m,属于北温带大陆性半干旱气候,年降水量270~280 mm,年平均气温7.4 ℃。植物物种约3000种,占新疆野生




注:底图采用自然资源部标准地图制作,审图号为GS(2020)4619号,对底图边界无修改。下同。

图1 研究区示意图

Fig. 1 Sketch map of the study area

植物总数的85%,主要树种为雪岭云杉(*Picea schrenkiana*)、桦树(*Betula*)、花楸(*Sorbus pohuashanensis*)等,其中雪岭云杉作为天山山地最主要的地带性森林植被,是构成天山乃至新疆森林生态系统的物质主体[24-25],天山林区内90%以上的林地均有雪岭云杉生长。本研究选择天然雪岭云杉为主要研究对象,使用直升机载激光雷达扫描巩留县林区内面积约为90 km²的研究区域,同时布设30个实测数据样方(25 m×25 m)。

### 1.2 技术路线图

研究技术路线主要由四部分组成(图2),第一部分为样方实测获取调查,获取样方内天然云杉的单木树高以及胸径,构建其回归方程;第二部分为机载激光雷达获取的点云数据,经过点云预处理步骤后,进行单木识别,获得整个研究区范围内的单木基础结构参数;第三部分为GEDI数据的获取、筛选以及与机载数据叠加构成多元融合数据框架,同时计算GEDI光斑内的地上生物量大小;第四部分主要以GEDI 81个非负RH值为输入的特征变量,对应地上生物量为输出变量,使用深度学习算法,验证GEDI数据在估算地上生物量方面的可行性。

### 1.3 数据收集与处理

1.3.1 样方实测数据 在研究区域内,根据GEDI光斑位置分布、不同林分以及不同海拔高度等因素,以天然云杉林为研究对象,选取30个(25 m×25 m)样方对其内胸径大于5 cm的天然云杉进行每木检尺。此外,30个样方均使用PVC管、测绳进行区域划分,以差分GPS测量样方四至坐标,选择东北角

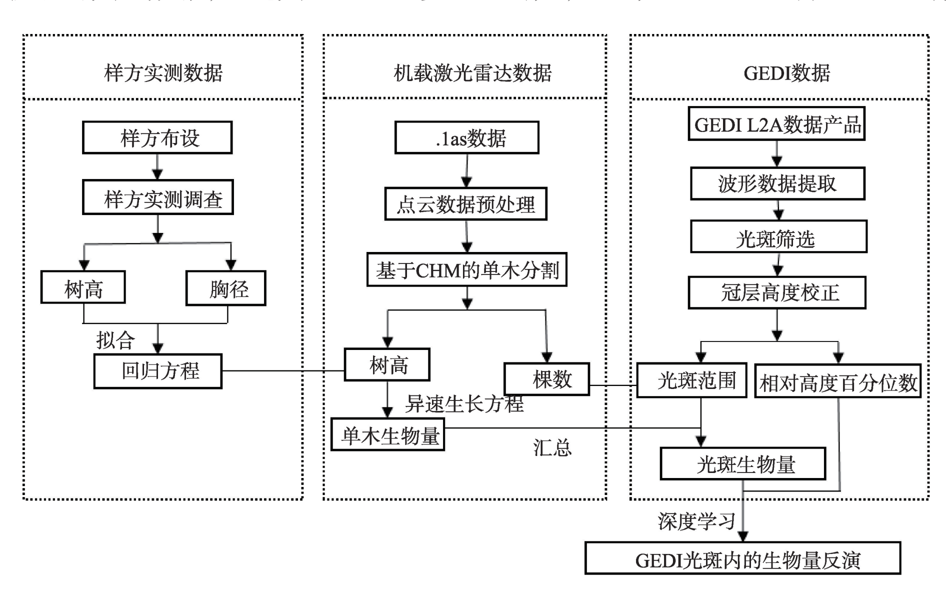



图2 技术路线

Fig. 2 Technology roadmap

为起点,用皮尺测量被测树到任意两边界线的交点距离,记录被测树的相对坐标,同时使用测高仪和胸径尺获得被测树的树高、胸径等参数。最终30个样方共测天然云杉单木686棵,获取其树高、胸径、相对位置等基础信息,根据样方内单木树高、胸径分布情况构建非线性回归关系(图3),最终生成的拟合公式可供胸径反演。

1.3.2 机载激光雷达点云数据

(1) 数据收集

飞行前,将实地地形与Google Earth卫星地图做详细对比,合理选择飞行区域,确保飞行安全。采用直升机搭载飞思IXU-RS 1000航摄仪、机载激光雷达(OPTECH Eclipse)扫描仪及GPS辅助航摄系统进行数据采集,获得研究区正射影像图和点云数据,飞行速度为4.8 m·s$^{-1}$,航带宽度为80~100 m,飞行器选择B-723U直升机,航向与旁向重叠率分别为60%、30%。需注意航摄结束后飞机安全着陆,航摄组立即获取影像和曝光点数据并检查质量,须补摄或重摄时选择合适的天气执行航摄,最终生成*.las文件和以供后期数据检查校准的影像数据。

(2) 数据处理

机载LiDAR点云数据的处理主要包括前期点云数据预处理阶段和后期数字产品生成与分析两个部分,此次研究使用LiDAR360(http://www.lidar360.com/)软件实现点云数据的拼接、滤波、地面点分类、单木分割等工程化预处理操作,图4为直升机机载激光雷达点云数据处理流程,其中前期处理后的点云数据,经过插值方法生成基础数字地形产

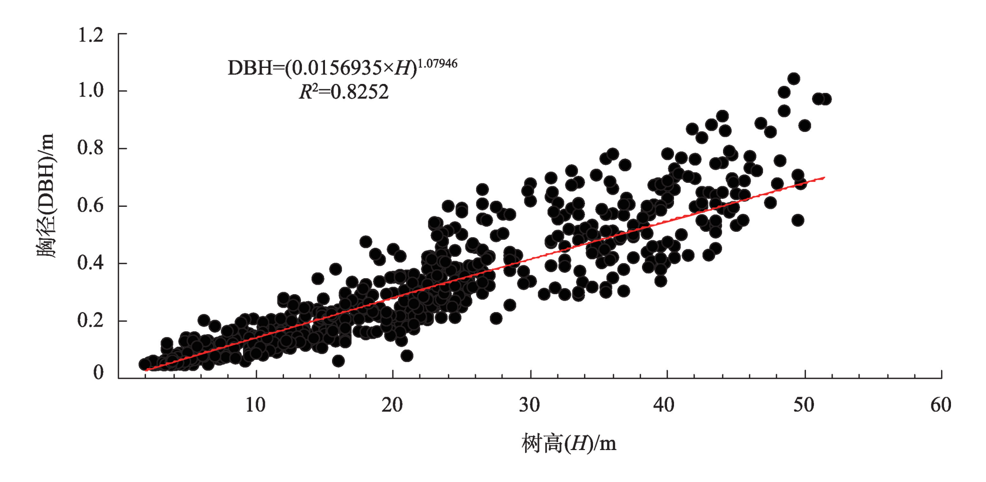



图3 样方单木树高-胸径回归拟合

Fig. 3 Regression fitting of $H$ and DBH in the sample plot

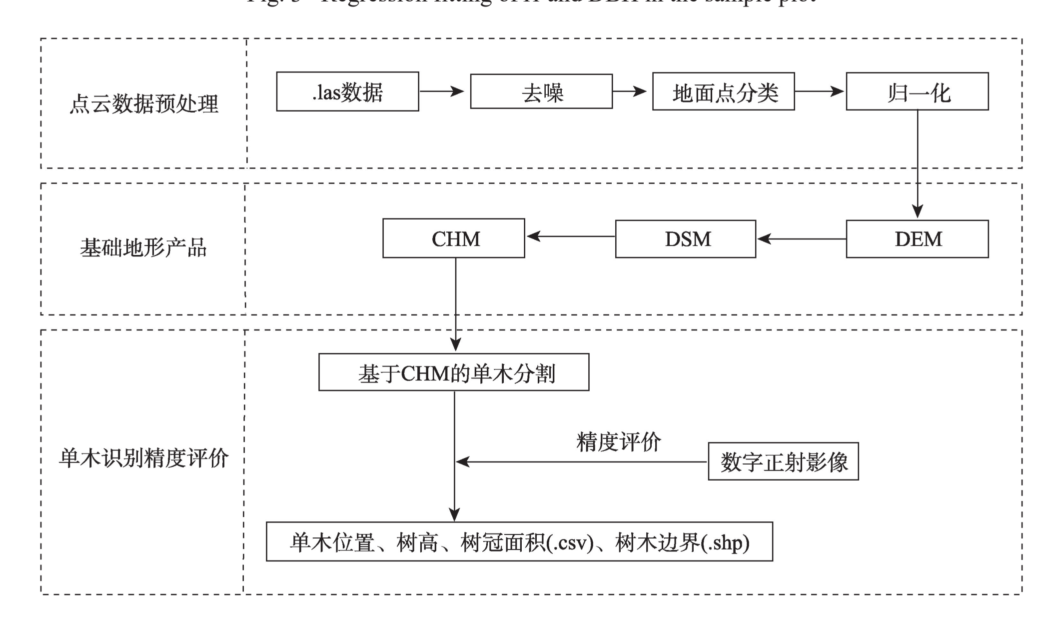



图4 直升机机载点云数据处理流程

Fig. 4 Flow chart of helicopter airborne point cloud data processing

品，包括数字高程模型（Digital Elevation Model，DEM）、数字表面模型（Digital Surface Mode，DSM）与冠层高度模型（Canopy Height Model，CHM）。

针对单木识别问题，研究选择基于CHM的单木分割方法，其本质是针对灰度图像做分割处理，最常用的算法是分水岭分割，即在生成的CHM影像中使用浸没模拟思想（immersion simulation）[26]，向相邻的山丘或高低分割开来，从局部最小值注入流水，将注入区域注满，根据图像的灰度变化建立屏障作为分割线，防止相邻区域注水相遇，最终所有区域均注满且无相互影响从而形成代表单个对象存在的“集水盆”，从而实现对单个物体的识别功能。图5为研究区局部单木识别效果图。

1.3.3 *GEDI数据* GEDI共由一个全覆盖激光器和两个全功率激光器组成，共发射四条波束，每个光束在光路上进行光学抖动，创建8个地面轨道，轨道间距约为600 m，共约4.2 km轨迹总宽幅，每条轨道上由多个相距60 m，直径约为25 m的圆形光斑组成，因此，沿轨道覆盖几乎连续[23,27]。同时，根据数据处理阶段的不同，GEDI数据产品主要分为4个等级（https://gedi.umd.edu/）。

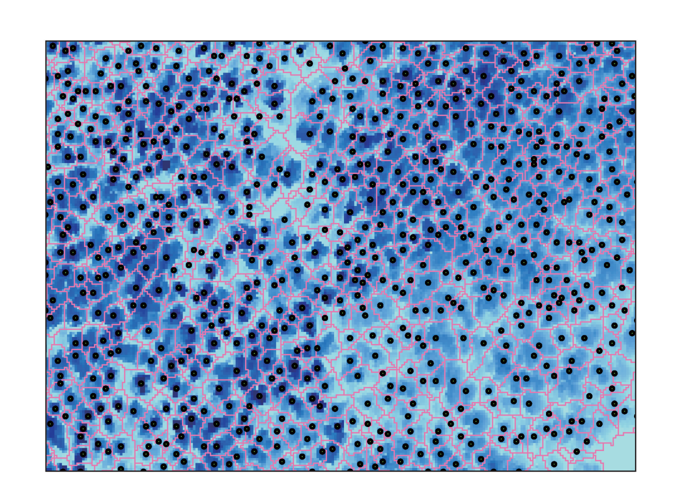


图5 基于CHM的单木识别

Fig. 5 Segmentation process based on CHM

（1）GEDI参数选择

本研究主要选择GEDI的L2A级数据产品，使用数据属性为相对高度百分位数RH，其中RH的获取主要依靠对其接收到的波形数据的分析与处理，取值范围为0~100，$RH_0$为最低可探测回波，确定地面位置，$RH_{100}$则为最高可探测回波，为冠层顶部，表示探测位置相对于地面位置的积累能量的100%处的高度。但并非所有波形数据均能满足森林资源调查，因此在数据筛选参数中设置相应条件[28]：

① quality_flag=1，光斑满足基于能量、灵敏度、振幅和实时表面跟踪质量的标准，质量良好，为有效波形数据。

② sensitivity>0.95，地面回波面积所占总回波面积的比例与1的差值大于0.90。

③ degrade_flag=0，数据处于“降级”时，形成最佳估计，数据性能良好。

可以发现，RH主要代表对应探测到的回波高度与光斑内平均地面高程之间的高差，所以某些探测到的高度会低于平均地面高度，显示负值状态，因此，本研究使用的GEDI基础数据为非负RH值（$RH_{20}$~$RH_{100}$，间隔1%）。

（2）构建多源融合数据框架

经过上述数据筛选，共从8264个GEDI光斑中选择1082个GEDI光斑用于本次研究，并将光斑内的所有单木AGB累加作为GEDI光斑的实测AGB，用于后期深度学习算法建模。其中研究区涉及的GEDI光斑（直径约为25 m的蓝色圆形光斑）和光斑内的单木分布状态（绿色波点）如图6所示。

机载激光雷达由于自上而下的扫描方式，难以获得森林冠层下方的三维结构信息，故通过30个样方实测数据构建树高、胸径回归方程，根据机载LiDAR获得的单木树高求得其胸径，后求单木[29]和GEDI光斑内的AGB大小，如下式所示：

$$AGB_i = 0.1661DBH^{2.2528} \quad (1)$$

$$AGB = \sum_{i=1}^{n} AGB_i \quad (2)$$

### 1.4 深度学习

深度学习（Deep Learning，DL）作为机器学习的重要发展，突破传统机器学习（Machine Learning，ML）在数据预处理、算法选择以及调参困难等方面的限制。本研究利用Python实现基于AutoKeras框架的深度学习方法，选择StructuredDataRegressor模型，对GEDI的81个非负RH值（$RH_{20}$~$RH_{100}$）与光斑内的生物量构建模型来预测关系。

1.4.1 *模型反演* 对于选取的1082个光斑数据，使用Python语句实现StructuredDataRegressor模型，随机选择70%作为训练数据集（757个GEDI光斑），30%为验证集（325个GEDI光斑），共遍历40种模型，分别进行300叠加从而创建新的特征集，直至模

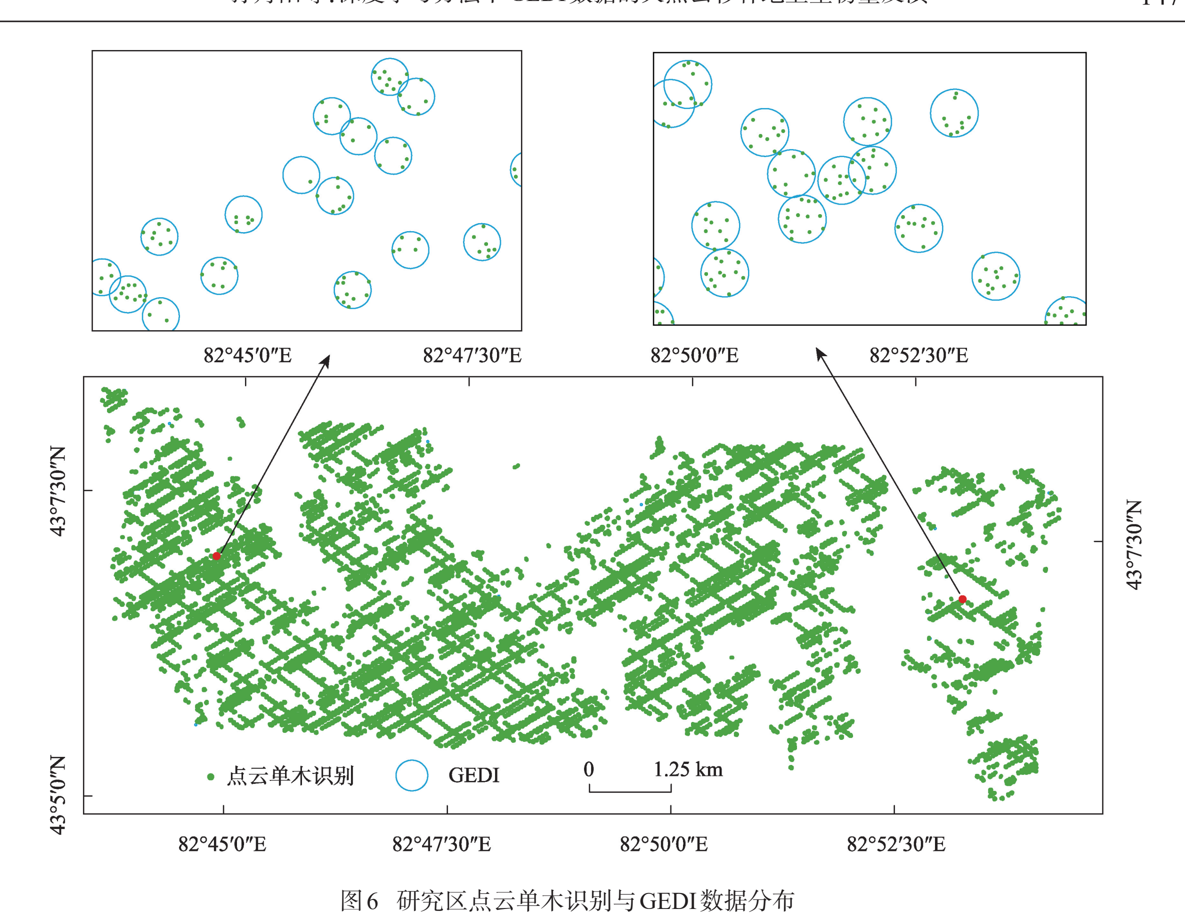



图6 研究区点云单木识别与GEDI数据分布

Fig. 6 Point cloud data individual tree recognition and GEDI data distribution in the study area

型到达预期效果,选择平均绝对误差(Mean Absolute Error,MAE)和损失函数Loss两项评价指标来衡量模型预测结果,MAE和Loss随着迭代次数的增加而变化(图7),同时随着迭代次数的增加,MAE和Loss逐渐降低,走势逐渐趋至平缓,生成的最佳模型MAE为3.73 mg·hm$^{-2}$,Loss为22.24。

深度学习算法生成的模型基础参数信息如表1所示,None是指将所有数据一次性放入模型进行训练,其后数字代表一维数据个数,即多少个变量($RH_{20}$~$RH_{100}$的81个特征变量),激活函数主要将非线性特征引入模型,优点在于计算简单、导数简单、收敛较快,其次模型的参数个数,等于上层神经元

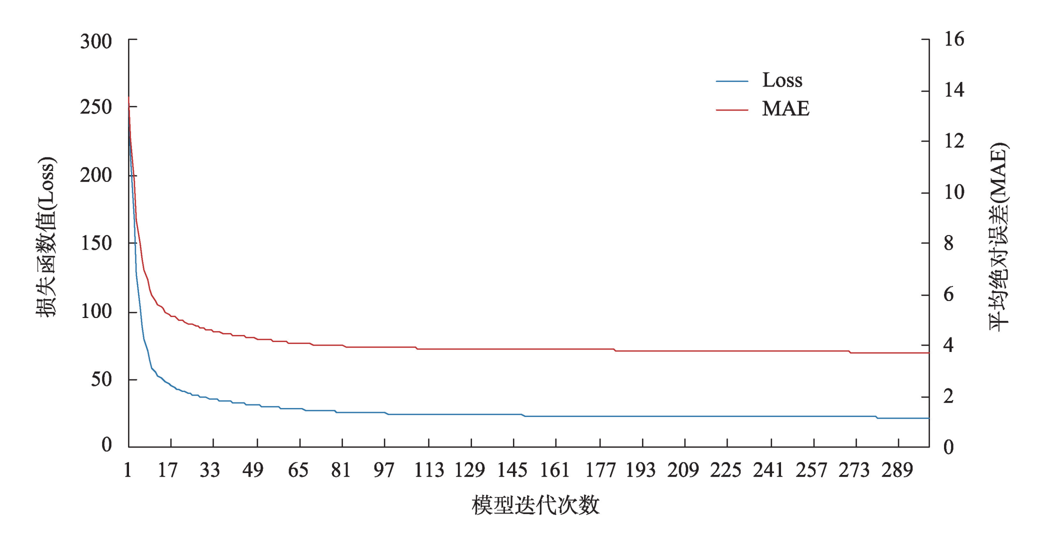



图7 深度学习模型MAE与Loss变化曲线

Fig. 7 Curve chart of MAE and Loss changes in deep learning model

表1 模型结构体系

Tab. 1 Model structure syst

| 网络层 | 输出大小 | 参数个数 |
|---|---|---|
| 输入 | [(None,81)] | 0 |
| 多类别编码层 | (None,81) | 0 |
| 归一化层 | (None,81) | 163 |
| 全连接层 | (None,64) | 5248 |
| 激活函数1 | (None,64) | 0 |
| 全连接层 | (None,1024) | 66560 |
| 激活函数2 | (None,1024) | 0 |
| 回归层 | (None,1) | 1025 |

个数×本层神经元个数,再加本层神经元个数。其中,参数总数为72996,参加训练的参数个数为72833,未参加训练的参数个数为163。

深度学习模型对训练集和验证集的预测情况由图8所示,可以发现,两个数据集的预测精度均有较好结果,AGB小于5 mg·hm$^{-2}$的数据被明显高估,极个别较大点被低估,大部分数据预测精度较高,误差控制在[-3,3]之间,这对于AGB主要分布于[10,30]的数值来说,影响较小,实际与预测数据之间有明显的回归趋势,呈较好的对应关系,通过数据统计发现,训练集、样本集以及整体数据的$R^2$分别为0.69、0.63和0.67,RMSE分别为4.68 mg·hm$^{-2}$、5.31 mg·hm$^{-2}$和4.88 mg·hm$^{-2}$,MAE分别为3.73 mg·hm$^{-2}$、4.22 mg·hm$^{-2}$和3.89 mg·hm$^{-2}$。

## 2 结果与分析

### 2.1 单木识别

根据公式(2)可知,GEDI光斑地上生物量包括光斑内所有单木AGB之和,因此在进行GEDI数据的AGB反演之前,需要对机载激光雷达技术的单木识别能力进行评价。此处选择召回率(Recal,*R*)、准确度(Precision,*P*)和代表分割整体精度的*F*值(*F*-

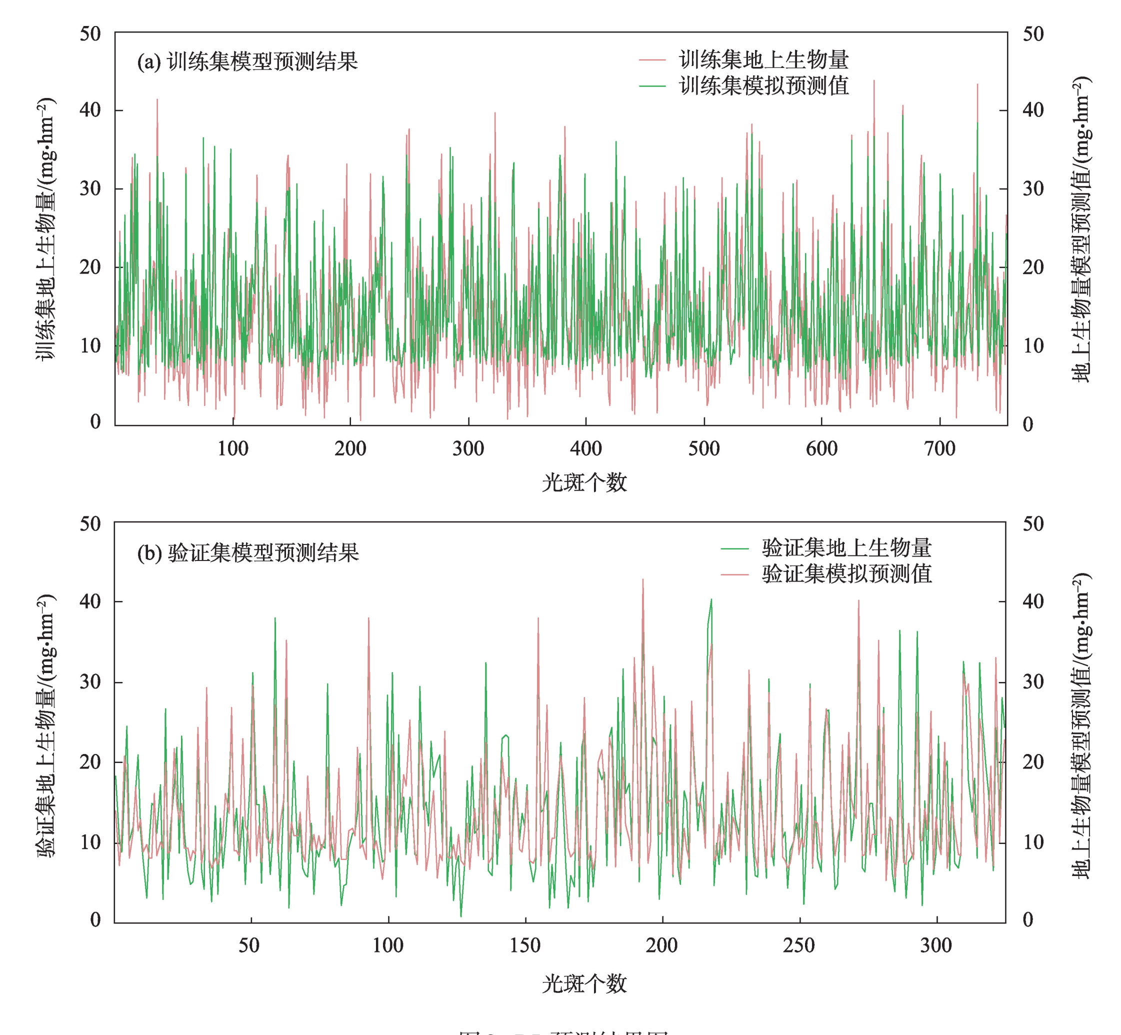



图8 DL预测结果图

Fig. 8 DL prediction result graph

score)[30]为评价指标,如公式(3)~公式(5)所示,其中 $TP$(True Positive)为最终正确提取的单木个数,$FP$ 为错误提取的单木个数(False Positive,$FP$),$FN$ 为未被提取的数量(False Negative,$FN$)。

$$R=\frac{TP}{TP+FN} \tag{3}$$

$$P=\frac{TP}{TP+FP} \tag{4}$$

$$F=\frac{2\times R\times P}{R+P} \tag{5}$$

结合正射影像图,在整个研究区随机选择3个云杉密度有所差异的样地(黄色标点)分别从R、P、F三个指标进行精度评价,图中红色标点为点云数据经处理后的单木识别情况(图9),3个样地的平均召回率为78.46%,准确度为88.16%,$F$ 值为82.5%(表2),对比发现,树冠之间高程起伏变化较大的单木正确识别率较高,高程变化不明显的单木会出现过分割和欠分割情况,说明机载LiDAR对单木识别具有明显可行性,可以准确获取树木空间坐标及各项单木结构参数,为生物量、碳储量的估算提供可靠数据[31]。

## 2.2 地上生物量误差分析

采用自动化深度学习算法实现GEDI光斑与多个非负RH之间的回归预测(表3),训练集、验证集、整体数据集的决定系数 $R^2$ 分别为0.69、0.63和0.67,均大于0.6,有较稳定的预测结果,RMSE(4.68 mg·hm$^{-2}$,5.31 mg·hm$^{-2}$,4.88 mg·hm$^{-2}$)和MAE(3.73 mg·hm$^{-2}$、4.22 mg·hm$^{-2}$、3.89 mg·hm$^{-2}$)数值较小,形成一种较稳定且高效方法,同时为验证该方法的适用性,对于最终预测得到的光斑AGB,选择反距离加权插值法(Inverse Distance Weight,IDW)绘制研究区地上生物量分布图(图10)。

可以发现,经插值形成栅格数据后,研究区地上生物量主要分布于7.25~39.58 mg·hm$^{-2}$,AGB的最大值出现在研究区的东北部区域,其次为西北方与南方均出现了局部峰值,整个研究区的AGB由南到北呈现逐渐递减关系,这主要是由于北方地区为草甸草原,AGB分布均匀,且明显小于云杉林的AGB数值大小。结合数字高程模型,发现单木生长主要集中于海拔1455~1900 m,坡度在9°~26°,2000 m海拔以上的区域单木数量减少,同时AGB分布稀疏,

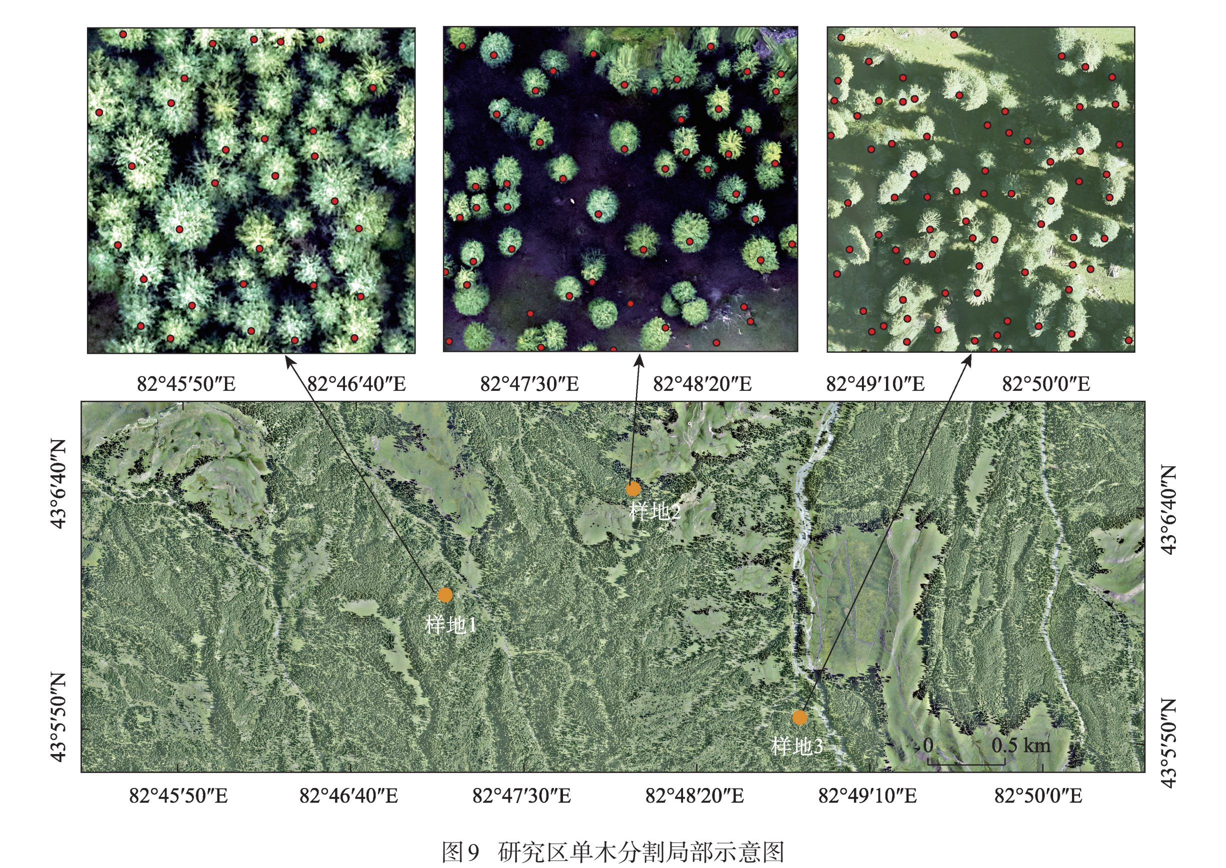



图9 研究区单木分割局部示意图

Fig. 9 Partial display of individual tree segmentation effect in the study area

表2 样地单木分割评价指标

Tab. 2 Evaluation index for individual tree segmentation of sample land

| 评价指标 | 召回率(*R*)/% | 准确度(*P*)/% | *F*值/% |
|---|---|---|---|
| 样地1 | 67.40 | 93.93 | 78.48 |
| 样地2 | 78.57 | 96.70 | 82.24 |
| 样地3 | 89.40 | 84.29 | 86.77 |
| 平均值 | 78.46 | 88.16 | 82.50 |

表3 深度学习下AGB的预测情况指标

Tab. 3 Prediction indicators for AGB under deep learnin

| 模型评价指标 | 数据集 | $R^2$ | RMSE /(mg·hm$^{-2}$) | MAE /(mg·hm$^{-2}$) |
|---|---|---|---|---|
| 深度学习(DL) | 训练集 | 0.69 | 4.68 | 3.73 |
| | 验证集 | 0.63 | 5.31 | 4.22 |
| | 总体数据 | 0.67 | 4.88 | 3.89 |

具体体现在研究区的最南部和最东部区域,该区域也位于整个研究区的主干河流位置,在同等条件下人为活动因素对其影响较大,因此会出现局部地区蜂窝状峰值情况。由此可见,本次研究在验证GEDI数据反演AGB能力的同时,为区域性乃至更大范围的地上生物量制图提供新的思路与方法。

分析整个数据处理过程,诸多因素尚未考虑,从而造成最终估算AGB存在误差,主要包括三个方面:(1) 样方实测数据易受人为主观影响,尤其是较高单木的树高测量方面,同时经点云数据获得的整个研究区单木参数也存在一定误差。(2) 整个研究区的单木识别普遍存在过分割和欠分割现象。(3) 点云数据基于CHM进行单木识别,通过对树冠的提取确定单木大概位置,但该位置与实际情况存在一定的空间不匹配情况,而地上生物量主要集中在枝干部分,树冠伸入或伸出样方范围均会对生物量计算造成一定影响,未考虑边界效应。

需注意的是,本次研究使用GEDI的$RH_{20}$~$RH_{100}$作为输入变量的主要原因,是由于研究区的某些地形较为陡峭,RH大都为负值,为统一研究基础,所以研究中使用的GEDI RH值均为正值。该做法使GEDI数据在输入变量个数方面有了大幅提升,尽可能挖掘了RH的基础属性价值,对于地形平缓地区,RH的取值可以更多。

最后,多源数据融合框架模拟生物量、蓄积量、材积等问题越发成为当下研究热点,GEDI、ICESat-2、Sentinel-1、Landsat为主要应用星载数据,结合其他Lidar技术,形成“天-地-空”一体的多模态数据框架,该方法对于全球性的森林资源清查、资源可持续管理具有一定的指导作用和技术论证,同时输入变量中增加高光谱影像数据,结合植被指数、地形坡度、纹理特征、降水等数据,提高回归模型的拟合精度,根据大量研究证实[18-19,32-34],该方法存在较大可行性。

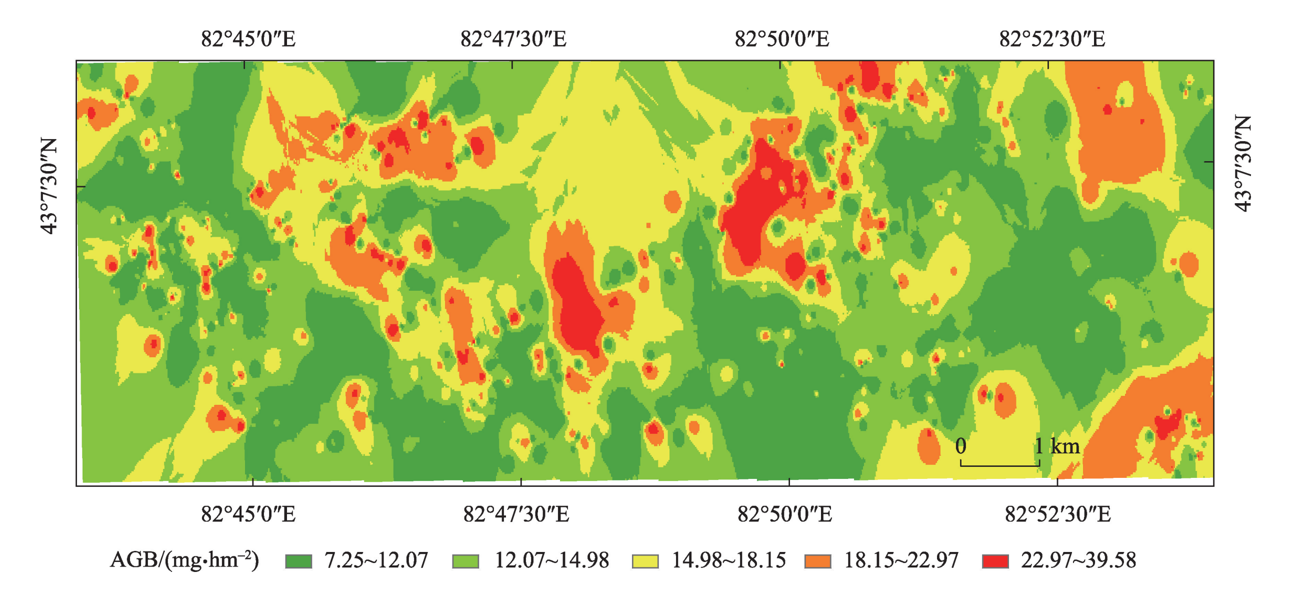



图10 研究区地上生物量分布示意图

Fig. 10 Schematic diagram of AGB in the study area

## 3 结 论

本文以新疆巩留县林区为研究区,构建多源数据融合框架,结合直升机机载LiDAR与星载GEDI数据,运用深度学习方法,建立GEDI多个RH值与光斑内地上生物量的预测模型,最终以实际计算地上生物量数据为基础,验证模型拟合精度,得出如下结论:

(1)直升机机载LiDAR获取大范围单木结构参数具有明显优势,目前已经形成了统一的作业规范,研究区平均单木识别率可达75%以上。

(2)GEDI光斑内生物量与81个非负的相对高程百分位数RH值使用AutoKeras框架下的深度学习算法进行模型预测,训练集、验证集以及整体数据集的最终模拟精度均在0.6以上,对最终输出变量有较高的解释程度,同时自动化深度学习实现了模型可视化,在一定程度上减少了深度学习作为机器学习"黑盒算法"的缺点。

目前深度学习方法在林业问题调查中显示出强大潜力,可以结合多源数据完成内部的多尺度学习,且无需显示特征的提取步骤,尤其是DL的非线性表达,可以高精度逼近复杂的非线性映射,使得最终输出模型具有强大的泛化能力。总之,多源数据融合,在基础数据精度得以保证的情况下,深度学习方法可以实现森林地上生物量的科学计算,也为森林资源调查提供一种思路。

**参考文献(References):**


[1] Olson, Jerry S, Watts J A, et al. Carbon in Live Vegetation of Major World Ecosystems[M]. Teen: ORNL-5862, 1983: 15-25.

[2] Hua Dengxin, Song Xiaoquan. Advances in lidar remote sensing techniques[J]. Infrared and Laser Engineering, 2008, 37(1): 21-27.

[3] 庞勇, 李增元, 陈尔学, 等. 激光雷达技术及其在林业上的应用[J]. 林业科学, 2005, 41(3): 129-136. [Pang Yong, Li Zengyuan, Chen Erxue, et al. Lidar remote sensing technology and its application in forestry[J]. Scientia Silvae Sinicae, 2005, 41(3): 129-136. ]

[4] 李增元, 刘清旺, 庞勇. 激光雷达森林参数反演研究进展[J]. 遥感学报, 2016, 20(5): 1138-1150. [Li Zengyuan, Liu Qingwang, Pang Yong. Review on forest parameters inversion using LiDAR [J]. Journal of Remote Sensing, 2016, 20(5): 1138-1150. ]

[5] 胡国军, 方勇, 张丽. 星载激光雷达的发展与测绘应用前景分析[J]. 测绘技术装备, 2015, 17(2): 34-37. [Hu Guojun, Fang Yong, Zhang Li. Development of spaceborne lidar and prospect analysis of surveying and mapping applications[J]. Geomatics Technology and Equipment, 2015, 17(2): 34-37. ]

[6] 谢栋平, 李国元, 赵严铭, 等. 美国GEDI天基激光测高系统及其应用[J]. 国际太空, 2018, 40(12): 39-44. [Xie Dongping, Li Guoyuan, Zhao Yanming, et al. GEDI space-based laser altimeter system and its application in the United State[J]. Space International, 2018, 40(12): 39-44. ]

[7] Adam M, Urbazaev M, Dubois C, et al. Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: Influence of environmental and acquisition parameters[J]. Remote Sensing, 2020, 12(23): 3948.

[8] 韩明辉. 基于星载激光雷达GEDI数据反演森林结构参数的研究[D]. 哈尔滨: 东北林业大学, 2022. [Han Minghui. Study on Forest Structure Parameters Inversion based on GEDI Data[D]. Harbin: Northeast Forestry University, 2022. ]

[9] Alireza H, Cheikh M, Annika K, et al. Deep learning for forest inventory and planning: A critical review on the remote sensing approaches so far and prospects for further applications[J]. Forestry: An International Journal of Forest Research, 2022, 95(4): 451-465.

[10] Cao Lin, She Guanghui, Dai Jinsong, et al. Status and prospects of the lidar-based forest biomass estimation[J]. Journal of Nanjing Forestry University, 2013, 56(3): 163-169.

[11] Torre-Tojal L, Bastarrika A, Boyano A, et al. Above-ground biomass estimation from lidar data using random forest algorithms[J]. Journal of Computational Science, 2022, 58: 101517.

[12] Quirós E, Polo M E, Fragoso-Campón L. GEDI elevation accuracy assessment: A case study of southwest Spain[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5285-5299.

[13] Fayad I, Baghdadi N N, Alvares C A, et al. Assessment of Gedi's lidar data for the estimation of canopy heights and wood volume of eucalyptus plantations in Brazil[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 7095-7110.

[14] Fayad I, Baghdadi N N, Alvares C A, et al. Terrain slope effect on forest height and wood volume estimation from GEDI data[J]. Remote Sensing, 2021, 13(11): 2136.

[15] Healey S P, Yang Z, Gorelick N, et al. Highly local model calibration with a new GEDI lidar asset on Google Earth Engine reduces Landsat forest height signal saturation[J]. Remote Sensing, 2020, 12(17): 2840.

[16] Fayad I, Baghdadi N, Bailly J S, et al. Analysis of GEDI elevation data accuracy for inland waterbodies altimetry[J]. Remote Sensing, 2020, 12(17): 2714.

[17] Potapov P, Li X, Hernandez-Serna A, et al. Mapping global forest canopy height through integration of GEDI and Landsat data[J]. Remote Sensing of Environment, 2020, 253: 112165.

[18] 林晓娟. 基于ICESat-2和GEDI森林冠层高度和森林地上生物量遥感诊断[D]. 北京: 中国科学院大学, 2021. [Lin Xiaojuan. Remote Sensing Diagnosis of Forest Canopy Height and Forest Aboveground Biomass based on ICESat-2 and GEDI[D]. Beijing: University of Chinese Academy of Sciences, 2021. ]

[19] 朱笑笑. 基于ICESat-2和GEDI数据的中国30米分辨率森林高度反演研究[D]. 北京: 中国科学院大学, 2021. [Zhu Xiaoxiao. Forest Height Retrieval of China with a Resolution of 30 m Using ICESat-2 and GEDI Data[D]. Beijing: University of Chinese Academy of Sciences, 2021. ]

[20] Ngo Y N, Ho Tong Minh D, Baghdadi N, et al. Tropical forest top height by GEDI: From sparse coverage to continuous data[J].Remote Sensing, 2023, 15(4): 975.

[21] Schneider F D, Ferraz A, Hancock S, et al. Towards mapping the diversity of canopy structure from space with GEDI[J]. Environmental Research Letters, 2020, 15(11): 115006.

[22] Roy D P, Kashongwe H B, Armston J. The Impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring[J]. Science of Remote Sensing, 2021, 4: 100024.

[23] Bauer L, Knapp N, Fischer R. Mapping Amazon forest productivity by fusing GEDI lidar waveforms with an individual-based forest model[J]. Remote Sensing, 2021, 13(22): 4540.

[24] 李明辉, 何风华, 刘云, 等. 天山云杉种群空间格局与动态[J]. 生态学报, 2005, 25(5): 1000-1006. [Li Minghui, He Fenghua, Liu Yun, et al. Spatial distribution pattern of tree individuals in the Schrenk spruce forest, Northwest China[J]. Acta Ecologica Sinica, 2005, 25(5): 1000-1006. ]

[25] 王雅佩, 王振锡, 刘梦婷, 等. 基于无人机影像天山云杉林主伐迹地提取研究[J]. 新疆农业科学, 2019, 56(7): 1312-1324. [Wang Yapei, Wang Zhenxi, Liu Mengting, et al. Research on extraction of final felling area of *Picea schrenkiana* var *tianshanica* based on UAV image[J]. Xinjiang Agricultural Sciences, 2019, 56 (7): 1312-1324. ]

[26] Soille P. Morphological Image Analysis- Principles and Applications[M]. Berlin: Springer-Verlag, 2003.

[27] Dubayah R, Hofton M, Blair M J B, et al. GEDI L2A elevation and height metrics data global footprint level V001[DB/OL]. NASA EOSDIS Land Processes Distributed Active Archive Center, 2023-08-15, https://doi.org/10.5067/GEDI/GEDI02_A.001.

[28] Dubayah R, Blair J B, Goetz S, et al. The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography[J]. Science of Remote Sensing, 2020, 1: 100002.

[29] 张绘芳, 高亚琪, 朱雅丽, 等. 新疆雪岭杉生物量模型对比研究[J]. 西北林学院学报, 2015, 30(6): 52-58. [Zhang Huifang, Gao Yaqi, Zhu Yali, et al. A comparative study on biomass models for *Picea schrenkiana* in Xinjiang[J]. Journal of Northwest Forestry University, 2015, 30(6): 52-58. ]

[30] Li Wenkai, Guo Qinghua, Jakubowski M K, et al. A new method for segmenting individual trees from the lidar point cloud[J].Photogrammetric Engineering and Remote Sensing, 2012, 78: 75-84.

[31] Hofton M, Blair J B, Story S, et al. Algorithm Theoretical Basis Document(ATBD) for GEDI transmit and receive waveform processing for L1 and L2 products[EB/OL]. Goddard Space Flight Centre, 2019. https://lpdaac.usgs.gov/documents/581/GEDI_WF_ATBD_v1.0.pdf.

[32] Rishmawi K, Huang C, Schleeweis K, et al. Integration of VIIRS observations with GEDI- lidar measurements to monitor forest structure dynamics from 2013 to 2020 across the conterminous United States[J]. Remote Sensing, 2022, 14(10): 2320.

[33] 胡涛. 基于多源数据的孟家岗林场针叶林蓄积量估测研究[D]. 哈尔滨: 东北林业大学, 2022. [Hu Tao. Estimation of Coniferous Forest Volume in Mengjiagang Forest Farm Based on Multi-source data[D]. Harbin: Northeast Forestry University, 2022. ]

[34] Knapp N, Huth A, Fischer R. Tree crowns cause border effects in area- based biomass estimations from remote sensing[J]. Remote Sensing, 2021, 13(8): 1592.

# Study on using deep learning method to retrieve the biomass of natural *Picea* forest from GEDI data

SUN Danyang[1,2], WEI Jianxin[1,2,3], YANG Liao[4], WANG Jie[5],
TANG Yuqi[4], Babierjiang DILIXIATI[2,3]

(1. College of Geography and Remote Sensing Sciences, Xinjiang University, Urumqi 830017, Xinjiang, China;
2. Xinjiang LiDAR Application Engineering Technology Research Center, Urumqi 830002, Xinjiang, China;
3. Xinjiang Uygur Autonomous Regions Natural Resources Information Center, Urumqi 830002, Xinjiang, China;
4. State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 5. School of Geographical Sciences, China West Normal University, Nanchong 637002, Sichuan, China)

**Abstract:** As the largest carbon reservoir on land, forests play a crucial role in human life and development. Understanding the dynamic changes in forest resources and modernizing their sustainable development is currently a significant research focus. This study focuses on natural *Picea* forests in the Tianshan Mountains and uses ground measurement data, helicopter airborne LiDAR point cloud data, and Global Ecosystem Dynamics Investigation (GEDI) data to construct a multisource fusion data framework. By utilizing deep learning algorithms within the AutoKeras framework, the study aims to predict the regression model of multiple relative height quantiles of GEDI data and their aboveground biomass in the study area, thereby validating the feasibility of GEDI data for large-scale aboveground biomass retrieval. The main conclusions are as follows: (1) GEDI data are highly feasible for estimating forest aboveground biomass. Through automated deep learning algorithms and training and verification sets, the overall data achieve a coefficient of determination ($R^2$) of 0.69, 0.63, and 0.67, respectively, along with a mean absolute error of 3.73 $mg \cdot hm^{-2}$, 4.22 $mg \cdot hm^{-2}$, and 3.89 $mg \cdot hm^{-2}$, demonstrating high prediction accuracy. (2) Helicopter LiDAR, an intermediate technology for estimating aboveground biomass using GEDI data, exhibits a single tree recognition accuracy of over 0.75 across the study area. The study successfully utilizes multimodal data fusion to quantitatively describe the structural parameters of the single tree foundation in the study area while verifying the potential of GEDI data for obtaining forest aboveground biomass. Moreover, the study provides a theoretical basis for estimating carbon sources and sinks, biomass, stock, forest management, biodiversity protection, and other projects in similar areas, offering essential guidance, and fundamental data support.

**Keywords:** natural *Picea* forest; GEDI; LiDAR; aboveground biomass; deep learning